Location Errors in Wireless Embedded Sensor Networks: Sources, Models, and Effects on Applications

Sasha Slijepcevic *sascha@cs.ucla.edu*
Seapahn Megerian *seapahn@cs.ucla.edu*
Miodrag Potkonjak *miodrag@cs.ucla.edu*
Computer Science Department, UCLA, Los Angeles, CA

Wireless sensor networks monitor the physical world by taking measurements of physical phenomena. Those measurements, and consequently the results computed from the measurements, may be significantly inaccurate. Therefore, in order to properly design and use wireless sensor networks, one must develop methods that take error sources, error propagation through optimization software, and ultimately their impact on applications, into consideration. In this paper, we focus on location discovery induced errors. We have selected location discovery as the object of our case study since essentially all sensor network computation and communication tasks are dependent on geographical node location data. First, we model the error in input parameters of the location discovery process. Then, we study the impact of errors on three selected applications: exposure, best- and worst-case coverage, and shortest path routing. Furthermore, we examine how the choice of a specific objective function optimized during the location discovery process impacts the errors in results of different applications.

I. Introduction

The impetus for research presented in this paper has two main sources: (i) the emergence of wireless embedded ad-hoc sensor networks (WEASN) and, (ii) the growing need for development of optimization algorithms that operate on noisy data and data with (potentially deliberately induced) errors. Wireless embedded ad-hoc sensor networks consist of a set of sensor nodes, each equipped with a number of sensors such as temperature, light, acoustic, seismic, and acceleration sensors. Nodes also contain digital communication systems, storage, processing resources, and in some cases actuators. Sensor nodes typically communicate using multi-hop schemes of ad-hoc wireless networks [19]. WEASNs have the potential to bridge the gap between the computational world of software and the Internet on one side and the physical, chemical, and biological worlds on the other [7]. One can easily envision numerous consumer, business, environmental, and scientific applications of WEASNs, ranging from early forest fire detection, indoor energy consumption monitoring, environmental monitoring [25], target tracking, and earthquake monitoring [7].

Accurate location information of each node is of crucial importance for organizing both communication and sensor-based measurements and calculations. Majority of the potential applications mentioned above, require a certain degree of knowledge of locations of nodes. Nodes could acquire the estimates of their locations from outside sources, where GPS [26] is one of the primary candidates for outdoor WEASNs. However, due to the large number of nodes in a sensor network, it is not justifiable from economic and energy preservation points of view to equip each node with a GPS receiver. The alternative solution proposed in [4, 15, 23, 22] is to design a location discovery algorithm that uses measurements of the distances between nodes and estimates of the locations of a small percentage of nodes (acquired through GPS, for example) to determine the locations of all or majority of nodes in a network. Both types of data, distance measurements and location estimates, are inherently noisy. Our goal in this paper is to show the importance of the estimation of the error in locations generated by location discovery algorithms.

The most obvious and maybe most important application of knowledge about errors and their behavior are development of localized optimization algorithms and corresponding objective functions. Applications that employ localized optimization algorithms can define a termination point and localized search boundaries based on the estimates of accuracy of the locations. Network management applications can use the estimate of error to detect intruder nodes that disperse invalid location data. Other applications include watermarking of solutions, equipment calibration, resiliency against skewing attacks, and reduction of the run time of algorithms without reduction of the quality of results. Additionally, accuracy of the location discovery process itself can improve if the error can be accurately modeled and estimated. Information acquired in different stages of the location discovery procedure can be selectively used according to error estimates, and the location discovery process can return confidence interval for its results and help applications to evaluate solution of the discovery process. In some instances, the estimate of error may indicate such a large error that renders any optimization effort ineffective. In such cases, as we show in Section VII.B, precomputed statistical information can be used instead of undertaking costly optimization steps.

Our discussions are organized around the following topics:

1. Modeling of the error in the initial distance measurements and location estimates: We define the sources of error in distance measurements and the initial locations of nodes, and we model the errors generated by those sources.

2. Measuring the quality of a solution generated by the location discovery procedure: There are situations where an assessment of the quality of a solution generated by the location discovery process may be necessary. We examine

methodologies of estimating the performances of the location discovery process and sensor networks application based on parameters of the network. Then, we show how the objective functions used in optimization-based location discovery procedure can predict performances of applications.

3. Effects of the location error on the performance of typical WEASN applications: The applications accept the results of the location discovery procedure as inputs. We examine how the error in inputs impacts the error in outputs. It is important to note that the knowledge about the effects of location errors helps not only the evaluation of the overall quality of service provided by a network, but also in guiding the design and resource management of WEASNs. We provide examples of applications that can determine the level of error in locations above which the results of the location discovery process are practically useless for that particular application. Therefore, the resources should not be consumed for the location discovery process under such circumstances.

I.A. Paper Organization

The remainder of the paper is organized as follows: In Section II, we provide a brief review of the related work. In Section III, we describe the preliminaries of the location discovery process, and we present an overview of five major error sources. An error model for distance measurements and initial locations is described in Section IV. Section V contains a proposed methodology of evaluation and estimation of the overall quality of the location discovery results. In Section VI, we briefly describe the three applications that we use to evaluate the impact of location errors on other WEASN tasks followed by Section VII with experimental results for both simulated and statistical data. We then highlight some key points of this paper in the conclusions.

II. Related Work

Wireless ad-hoc sensor networks in general and location discovery in particular, are intrinsically multidisciplinary topics. Therefore, a wide body of related knowledge and results exists in the literature. We focus our attention to wireless ad-hoc networks, radio-propagation models, techniques for location discovery in wireless environments, numerical analysis aspects, and discrete optimization techniques.

A number of applications are envisioned for WEASNs with the potential to dramatically alter human life [7]. However, a number of challenging technical problems associated with wireless ad-hoc networks and sensor networks must be solved before the sensor networks are widely used. The most important directions of research in WEASNs include new types of signal processing [28, 17], operating systems features [2], self-organization and deployment [3], low power design [19], integration, issues related to embedded systems [5], robotics [27], and integration with biological entities [1]. Location discovery in wireless sensor networks is also recognized as one of the basic tasks. Several location discovery systems have been developed recently [4, 18, 9, 30, 23, 22]. A common characteristic of location discovery algorithms in wireless networks is that they use distance measurements and initial location estimates as inputs. In [24, 31] the distances are measured using *Time Difference of Arrival* (TDOA). Two sources of error in TDOA long-range distance estimates are distinguished. The first source of error is the standard system measurement noise, modeled as a zero-mean random variable. The second type of error is the NLOS error, a product of the reflection of the signal when the line-of-sight path is obstructed. Since a reflected path is longer than the line-of-sight path, the NLOS error produces a positive bias in distance measurements. Another RF-based distance measurement technology estimates distances using *Received Signal Strength Indicator* (RSSI). It is used in an indoor user location and tracking system [4], as well as in [22]. In [4], a coverage map of the building lessens the impact of the sources of errors. Similar approach can be applied in mobile telephony, as proposed in [12]. However, for WEASNs deployed in an ad-hoc manner across a remote area, a profile of the area is likely not to be available.

Sources and characteristics of errors in wireless sensor networks are similar to those in other wireless networks, especially if the radio signal is used for the distance measurements. Besides radio signal, sound is identified in several projects as another candidate for ranging. In [23, 18] the distances are measured using the time difference between simultaneously transmitted radio and ultrasound signals. Simulation results given in [23] show that with ranging error and initial location error for nodes with GPS both simulated as a 20 mm white Gaussian noise, the estimated location of nodes are within 20 cm from the actual positions. It must be mentioned however that ultrasound has a shorter range and lower tolerance for obstacles in the path than radio signals, and is included in the design of a sensor node solely for the purposes of localization. Acoustic-based distance measurements are used in [8], where it is reported that the precision of ultrasound is in the sub-cm range, when the line of sight is not obstructed, and microphones and speakers are directed at each other. In [8], it is proposed that the ambiguities created from obstructed line of sight can be solved either by employing additional sensors, e.g. cameras to detect obstructed line of sight, or using localization algorithms that detect inconsistencies.

Once the distances are measured and initial locations are known, we can pose the location discovery problem as a system of nonlinear polynomial equations and apply one of the the standard numerical analysis techniques to solve it. There are a number of excellent textbooks that discuss techniques for solving systems of nonlinear equations within the context of numerical analysis [10, 20]. It is well known that the numerical stability of polynomial equations with errors, as in location discovery case, is often surprisingly low [10, 20], and that very small perturbations in very few coefficients can result in significantly different solution than the intended one. Furthermore, numerical errors often accumulate quickly [10] and invalidate the final result.

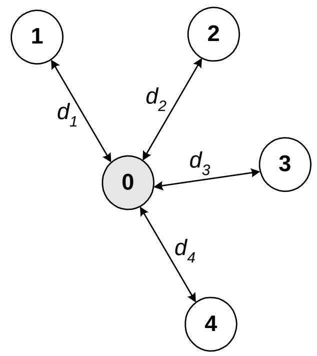

Figure 1: Multilateration Example

However, the exact analysis of error propagation is essentially intractable [29]. Therefore, we adopted the widely used practice in numerical analysis: statistical experimental evaluation of error propagation [20, 29].

III. Preliminaries

III.A. Location Discovery

Location discovery is a term used in WEASNs to describe the process by which the nodes determine their relative or absolute physical coordinates. In several existing schemes, this process typically consists of two basic steps: (i) distance measurements (ranging), and (ii) atomic multilaterations. Node locations can be determined from measured distances between nodes, and known locations of a subset of the nodes in a network, such as the base stations in wireless phone networks or GPS-equipped nodes in sensor networks. The distance measurement techniques proposed until now for use in sensor networks are:

1. *Radio signal strength* [22]: The distance between two nodes in a network can be estimated by comparing the transmitted power at the sender and the received power at the receiver of the message.

2. *Time of Arrival using radio signal and acoustic signal* [23, 8]: Since the propagation speeds of radio and acoustic signals are known, the distance between the sender and the receiver can be calculated from the time difference of arrival of two such signals.

After the ranging step, where the distances are measured between the nodes that can hear each other, the next step is the multilateration procedure. Multilateration combines the measured distances and the known locations of GPS-equipped nodes, or nodes which determined their locations in previous multilateration procedures, to determine the locations of other nodes. An example of multilateration is given in Figure III.A, where the node 0 can estimate its location based on information received from the nodes 1, 2, 3, and 4. We refer to the nodes that send estimates of their locations as *beacons*. To estimate its location, the node 0 computes a location for which the value of the chosen objective function is minimum. In our experiments in this paper, we use three well established objective functions:

$$L_1 = \sum_{i=1}^{n} |D_{i0} - R_{i0}| \tag{1}$$

$$L_2 = \sum_{i=1}^{n} (D_{i0} - R_{i0})^2 \tag{2}$$

$$L_\infty = \max_{i=1..N} \left| \frac{D_{i0} - R_{i0}}{R_{i0}} \right| \tag{3}$$

Here, R_{ij} is the estimated distance between the nodes *i* and *j*, D_{ij} is the distance between estimated locations of the nodes *i* and *j*, and *N* is the number of nodes that send their estimates to the node computing its location.

After the node 0 determines its location, it can become a beacon for other nodes. There are several algorithms proposed for propagation of estimated locations through the network, such as those presented in [22, 16, 15]. We adopt the iterative improvement based algorithm for the location discovery optimization process as presented in [15]. Although this technique has not been used in WEASN application domain, our choice was directed by successful use of iterative improvement algorithms for combinatorial optimizations [11] and continuous optimizations. Additionally, iterative improvement is well suited for localized and distributed implementation as detailed in [15].

III.B. Sources Of Error In WEASN

As mentioned in the introduction, we identify at least five main sources of error that influence optimization results in WEASNs:

1) Measurement
2) Finite precision
3) Objective function-specific
4) Intractable optimization tasks
5) Localized algorithms.

The first two, measurement and finite precision related errors are inherent in all physical computing systems. Measurement errors arise due to sensing technology limitations, phenomena instability, and environment noise. Numerous well studied techniques exist to reduce or compensate for such errors in many domains. Examples of such techniques are averaging methods that rely on several distinct measurements, and digital signal processing (DSP) techniques. Finite precision errors are the result of inaccuracies induced in the result due to limited computation precision of digital computers. Since the WEASN hardware are typically very resource constrained, such errors can be of significant importance.

The second two sources of error are optimization-task specific. The objective function used at the heart of the optimization process may not accurately or completely capture the essence of the problem and thus can lead to erroneous conclusions. The fourth source of error is due to the inherently intractable nature of the many optimization tasks tackled by WEASNs. Optimal solutions to intractable problems are by definition very difficult to compute. Furthermore, exceptionally limited resources such as energy, communication bandwidth, storage, and processing power make tackling such problems and their associated errors (in results) especially challenging in WEASNs.

As opposed to the sources of error described above, which exist in many computation domains, the error associated with localized algorithms is unique to distributed computation systems such as WEASNs. Localized algorithms are especially suited to WEASNs due to the geographic structure of the network. In localized algorithms, only nodes in spatial (or other) ”proximity” collaborate and participate in a computation. Using local information in computing inherently global metrics can be a very error prone task. Understanding such errors and their behavior is a fundamental task in designing successful localized algorithms.

IV. Error Modeling

Appropriate error models for distance measurements and GPS-generated locations are an important part of the design process for WEASN. Wireless sensor networks are intended to be used, among other environments, in remote and inhospitable areas, where the error characteristics of different measurements cannot be examined in advance. As we show in this paper, the knowledge of both, error sources and the error propagation through the stages of the location discovery process, can impact the design decisions at all levels.

For two technologies for distance measurement in sensor networks, RSSI and acoustic, described in Section III, we have developed separate error models taking into account the different sources and amplitudes of error specific for each technology. The RSSI error model is based on the path loss models in [21] and measurements from [23, 22, 31], while the error model for acoustic ranging is derived from measurements in [8, 23].

The main source of error in RSSI-based distance measurements is the complexity of modeling of environmental effects in the propagation model. Reflection, scattering, and diffraction, as well as antenna gains produce significantly different path losses for equal distances. For example, it is reported in [23] that placing a node 1.5 m above the ground increased the transmission range of the radio signal three times relative to when the node is on the ground, although in both cases there is a direct-line-of-sight path between the transmitter and the receiver. From [21] (pg. 104), the distribution of measured distances $\hat{d}$ for the correct distance *d*, is given as:

$$10n\log(\frac{\hat{d}}{d_0}) - 10n\log(\frac{d}{d_0}) = X_\sigma[dB], \quad (4)$$

where X_σ is a zero-mean Gaussian random variable with standard deviation σ, both given in dB. From Equation 4, the distribution of the RSSI error is:

$$RSSI\_ERROR(d) = \hat{d} - d = d(1 - 10^{\frac{X_\sigma}{10n}}) \quad (5)$$

RSSI_ERROR depends on the measured distance *d*. Standard deviation σ defines the one σ range as a percentage of the measured distance. For example, $\sigma = 1$ generates 68% of distance measurement errors within 6% of the measured distance, for the value of the environment constant $n = 4$. The dependence of the error on the distance *d* in the model causes larger errors for larger distances, which is consistent with the measurements from [23]. Finally, the simulated distance measurements are generated as:

$$R_{ij} = D_{ij} + RSSI\_ERROR$$

where R_{ij} is the measured distance between nodes *i* and *j*, D_{ij} is the correct distance, while *RSSI_ERROR* represents the RSSI error.

Acoustic ranging achieves much better accuracy than RSSI [23, 6]. Such results make acoustic ranging a preferred technology for distance measurements. Our modeling of the ranging error in acoustic ranging systems is based on the results reported in [8]. There are three important sources of error in acoustic ranging that cannot be eliminated by averaging distance measurements over time [8]:

1. Non-line-of-sight (NLOS) error: This error occurs when there is an obstacle between nodes. We model it as a distance dependent, uniformly distributed, positive error *NLOS=Un(0, d*NLOS_ERROR_MAX)*, where *d* is the measured distance, and the constant *NLOS_ERROR_MAX* belongs to the interval [0,1].

2. Speed of sound error: Atmospheric changes in the environment, as well as different atmospheric conditions in various parts of the network impact the speed of sound. We model this as a distance dependent Gaussian noise *N(0,*$\sigma_{SSE}(d)$*)*. The speed of sound error is given as *SSE=N(0,*$\sigma_{SSE}(d)$*)*.

3. Orientation error: The emitter and the sound sensor may not be aimed directed towards each other, which produces error that depends on the angle between them. We model this error as the angle-dependent Gaussian noise *OE=*α**N(0,*$\sigma_{OE}(d)$*)*, where α is the angle between the emitter and the sound sensor. Thus, the acoustic distance measurement between the nodes *i* and *j* is simulated as:

$$R_{ij} = D_{ij} + NLOS + SSE + OE$$

For each simulation, a subset of nodes that have initial estimates of their locations (beacons) is randomly selected. The initial locations of beacons are generated by superimposing an error to the real locations as follows. The real locations of sensor nodes A_i, *i=0,..,n* are represented as points $A_i(x_i, y_i)$. Coordinates x_i and y_i are generated from two uniform distributions, one on the interval $[0, X_{max}]$ and one on the interval $[0, Y_{max}]$. The error is generated from Rayleigh distribution, by generating a pair (Δx_i, Δy_i), where both Δx_i and Δy_i are selected from a zero-mean Gaussian distribution $N(0, \sigma_{loc})$. The average location error is then:

$$\mu = \sqrt{\frac{\pi}{2}}\sigma_{loc}. \quad (6)$$

Now, by selecting σ_{loc}, we can generate a location error distribution with the desired mean value.

V. Evaluation of Solution Quality

The output of a location discovery algorithm consists of the estimates of locations of the nodes in a network. There are two issues that we address here:

1. How does one evaluate the quality of a given solution in a simulation environment, where the real locations of nodes are known?

2. How does one estimate the quality of a solution in a working environment, where the real locations of nodes are not known?

Examples of applications that may require an estimate of the quality of a solution are given in Section VI. The notations used in the following definitions are as follows:

(x_i,y_i) - Real location of the sensor node *i*

(x_i^s,y_i^s) - Estimated location of the sensor node *i* in the solution *s*

d(n_i,n_j) - Measured distance between the nodes *i* and *j*

Two quality functions that evaluate different properties of a solution are:

1. Average location error:

$$QF_1(s) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_j^s - x_i)^2 + (y_j^s - y_i)^2}$$

2. Maximum location error:

$$QF_2(s) = \max_{i=1..N}\sqrt{(x_j^s - x_i)^2 + (y_j^s - y_i)^2}$$

The quality function QF_1 calculates the average location error, implying that all nodes are taken into account. QF_1 is aimed at applications whose performance depends on locations of majority of the nodes in the network. However, information about the average location error may not satisfy requirements of applications that rely on a subset of nodes in the network. Performances of such applications may depend more on a maximum location error in a solution. The quality function QF_2 measures this property of a solution.

Now, in order to answer our second question, we accept that the quality functions QF_1 and QF_2 represent valid assessments of the quality of a solution. Our task now is to establish functions that estimate the quality of a solution based on information available during location discovery process, i.e. the estimates of the locations of sensor nodes and the measured distances between them. Also, two estimation functions must correlate to the quality functions, so that applications can use either of the estimation functions, depending on what property of a solution is to be estimated. The evaluation functions that correspond to the quality functions QF_1 and QF_2 are:

1. Sum of distance inconsistencies:

$$EF_1(s) = \sum_{i=1}^{N}\sum_{j=1}^{i-1}\left|d(n_i,n_j) - \sqrt{(x_j^s - x_i^s)^2 + (y_j^s - y_i^s)^2}\right|$$

2. Maximum inconsistency:

$$EF_2(s) = \max_{\substack{i,j=1..N \\ i\neq j}}\left|d(n_i,n_j) - \sqrt{(x_j^s - x_i^s)^2 + (y_j^s - y_i^s)^2}\right|$$

The first estimation function EF_1 calculates the sum of the differences between the measured distance from the node *i* to the node *j*, and the distance between the estimated locations of the nodes *i* and *j* in the solution *s*, for each pair of nodes (*i*,*j*). EF_1 estimates the quality of a solution taking into account all nodes in a network and therefore corresponds to QF_1.

The second estimation function EF_2 estimates the quality of the solution *s* based on the largest difference between a measured distance and distance between estimated locations. The objective function EF_2 informs applications about the maximum location error that can be expected to occur in the solution by measuring the largest inconsistency between a measured distance and the distance between corresponding nodes . EF_2 corresponds to QF_2.

In the following experiments, we examine the degree of correlation between the evaluation functions and the quality functions. We generate random sensor network topologies with 50 sensor nodes placed in a 40m x 40m field. The transmission range of the nodes' radios is 10m, and we simulate the error in the locations of sensor nodes by adding random location errors to the real locations. The average location error, i.e. average distance between the real locations of nodes and the estimated locations from the solution *s*, is a parameter, and can be controlled as given in Equation

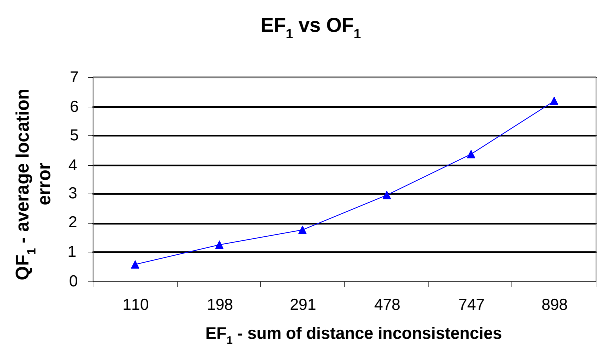

Figure 2: Correlation between the estimation function EF_1 and the average location error (QF_1).

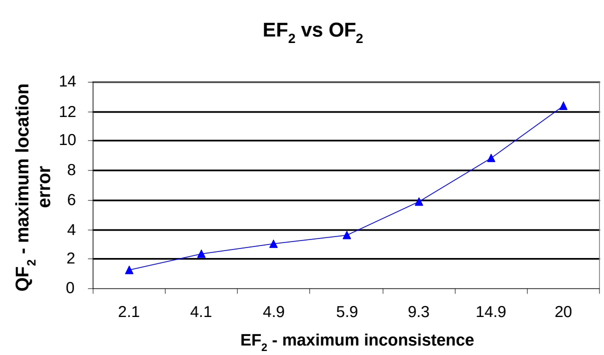

Figure 3: Correlation between the estimation function EF_2 and the maximum location error (QF_2).

6 in Section IV. For each value of the average location error, a set of 100 networks is generated. The average values for two quality and two estimation functions are calculated for each set. Figure 2 and Figure 3 show correlation between these values. The coefficient of correlation is 0.99 for both pairs of functions.

VI. Effects Of Location Errors On Applications

We now turn our attention to studying how the location errors discussed so far impact different applications that may rely on the faulty information. We have two main goals in our study: (i) to see whether estimating errors in locations can predict certain errors in the results of the algorithms which rely on the location information, and (ii) which estimation function performs better for a specific application. Clearly, a countless array of application algorithms exist that may rely on faulty data. Since WEASNs have been the focus of our case studies, the applications can include wireless network management and operation related tasks such as finding optimal routing paths and clustering, or sensor network specific tasks such as sensing, tracking, and coverage. Although one can expect that errors will effect each application differently, as mentioned before, we opt to focus our attention on the following three: (i) best-case (maximal support) and worst-case (maximal breach) coverage, (ii) minimal exposure, and (iii) shortest paths. As we will discuss, these three inherently different location dependent applications each have distinct properties that we believe are important in studying the performances of different location-error estimation functions. We discuss these specific properties along with some intuitive observations after presenting a brief introduction and background for each case. The experimental results provided in the subsequent section serve as a guide in studying what real impacts location errors may have on the results produced by the selected applications and whether certain predictable behavior patterns exist.

VI.A. Maximal Breach and Maximal Support

Recently, [13] presented the maximal breach and maximal support paths and algorithms for their calculation as a method for characterizing and computing the worst- and best-case sensor coverage in sensor networks. The maximal breach path there is defined as a path through the sensor network connecting a given starting point to a given end point such that the closest point on the path to any sensor is maximized. We refer to the closest distance to a sensor along the maximal breach path as *breach*. The significance of *breach* is that if an object is traveling through the sensor field, it must be within *breach* distance of a sensor, at least once, even if it optimally tries to avoid the sensors (be as far as possible).

Similarly, the maximal support path is defined as a path through the sensor network connecting a given starting point to a given end point such that the farthest point on the path to the closest sensor is minimized. We refer to the farthest distance to a sensor along the maximal support path as *support*. Analogously to *breach*, the significance of *support* is that if an object is traveling through the sensor field, it must be at least *support* distance away from the closest sensor, at least once.

Here, we are mainly interested to see how location errors affect the computation of the values of *breach* (worst-case coverage) and *support* (best-case coverage). For this application, typically the location of two nodes determine the final outcome of the computation. In the case of *breach*, this is where the maximal breach path is closest to sensors while for support, the nodes farthest from each other along the maximal support path are most relevant. As we will see below, here the location error of a small set of nodes may strongly influence the outcome, which is in strong contrast to the other two cases where the locations of a larger number of sensors potentially impact the results.

VI.B. Minimal Exposure Path

As presented in [14] exposure in a sensor networks can be used as a measure of how well the sensors observe moving objects in a field. In general, [14] defines the exposure for an object moving in the sensor field during an interval $[t_1, t_2]$ along the path *p(t)* as:

$$E(p(t), t_1, t_2) \triangleq \int_{t_1}^{t_2} I(p(t)) \left| \frac{\mathrm{d}(p(t))}{\mathrm{d}t} \right| \mathrm{d}t$$

where *I(p)* is a non-negative sensor field intensity function representing the sensor field strength at a point *p*. Hence, the *minimal exposure path* is a path connecting a given starting and ending point along which the exposure integral is minimized. The exposure of an object traveling along a path through the sensor field is one measure of the likelihood that the sensor nodes will detect and observe that object. Consequently, one can compare the sensor coverage provided by two different sensor network instances by computing the exposure along the minimal exposure path in each network, with the network with higher exposure along its minimal exposure path providing the higher coverage level.

Unlike in *breach* and *support* calculations where the final results ultimately depend on the locations of one or two sensors, the computation of the minimal exposure path is typically affected by the location of several (if not all) nodes in the network. The exposure along each segment of the path is potentially influenced by several sensors that are geographically close to the region. Thus one can expect that the overall error in locations of the nodes will influence the outcome of the calculations more significantly than the error in the location of a particular node or set of nodes. This effect can be more prevalent if systematic errors exist in the computed node locations since random errors may cancel each other.

VI.C. Shortest Paths

Geographically short paths connecting nodes have several interesting applications in WEASNs. For example, if we assume that node transmission ranges are fixed, then transmitting along the geographically shortest path connecting an arbitrary pair of nodes may be one method of reducing communication energy costs. Although finding routing strategies that minimize and/or balance the overall energy consumption is an interesting and challenging task, here we focus our attention on how location errors affect the computation of shortest paths connecting arbitrary pairs of nodes. To analyze this application, we assume that each node has a predefined communication range *d*. For every pair of nodes *i* and *j*, we define D_{ij} as the Euclidean distance between the nodes *i* and *j*, computed based on the given (possibly faulty) location information. We then build a weighted graph *G(V,E)* corresponding to the topology of the WEASN where for each pair of nodes $i \in V$ and $j \in V$ with $D_{ij} \leq d$ we add the edge e_{ij} in *E*. The weight of each edge e_{ij} is simply set as D_{ij}. We then compute the all-pairs-shortest paths in G. We measure the application error by computing the shortest paths using the known real locations and the faulty locations produced by the location discovery algorithm and counting the number of paths that are different in each case. Note that for *n* nodes, we have at most n^2 paths to consider. Clearly numerous other error metrics can be used to analyze this application. Indeed, selecting the appropriate application error model may significantly impact the way application performance is judged. However, our main focus here is to study how errors in one level impact the computations at another level so the specific choice of error models is less relevant.

One of the aspects that sets this application apart from the other two is that in essence, some nodes are more important than others when shortest paths are concerned. For example, in instances where nodes are uniformly distributed in a field, those in the middle of the field are clearly more likely to participate in shortest paths connecting arbitrary nodes in the network. Thus, the location errors of such nodes may play a more pivotal role in affecting the final results than the location errors in nodes that are relatively isolated or along the boundaries of the field.

VII. Application Results

In our experimental setup we create a random network topology for which we know the exact node positions. We refer to this data as *true* locations. The results of the location discovery process, referred to as *faulty* locations, are then processed by each application and compared to the results obtained using the true locations to compute the error metrics. The goal here is to compute the error in results using the available *faulty* location information and the results which would have been obtained if *true* location information were available. Unless specified otherwise, for each case, we deploy 50 nodes in a square field that is 40 meters wide. For all minimal exposure, *breach*, and *support* calculations we assume that an agent is traveling from the boundary line x=0 to the opposite boundary line x=40m. Note that the apparent clustering of data points in the following plots are the result of the granularity at which we varied the error levels in the simulations.

VII.A. Application Errors Caused by Faulty Location Data

In Figure 4 we determine the correlation between the location-error evaluation functions and the errors observed in the minimal exposure, breach, and support applications. We ran 2000 different simulations. For each case we estimated the location error using both evaluation functions described in section V. Then, for each case, we calculate the percentage difference between application results with the real locations and the results with faulty locations. The values for the quality functions are normalized. As noted on the figure, data points marked by a '.' correspond to values obtained using the maximum inconsistency (EF_2) and the data points marked by a '+' correspond to values obtained using the sum of inconsistencies (EF_1) estimation functions.

The top two plots in Figure 4 show the % error in the computed minimal exposure to cross the field relative to the minimal exposure along the path obtained using the true sensor locations using the all-sensor and the closest-sensor exposure models respectively. Since in practical instances, the true locations of the nodes will not be known (hence the need for a location discovery process), one must consider the impact of following optimization decisions based on faulty data. Consequently, the middle two plots in Figure 4 provide an alternative to measuring the errors reported in the first row. Here, the % error is computed by comparing the exposure along the minimal exposure path found using faulty locations to the exposure along the *same* path using the true node locations. The final row of plots show the *breach* and the *support* errors. For both these cases, the errors are computed by comparing the *breach* and *support* obtained using the faulty data with those computed using the true locations. In all plots in Figure 4, the solid line represents the linear regression of the '.' data set while the dotted line is the regression line corresponding to the '+' data set.

The data in Figure 4 shows the correlation trends between the reported objective functions and the observed errors. Although at a fine grained level this correlation is not obvious, the overall picture is similar for all the cases. In almost all cases the regression line has a smaller slope for the sum data set which may indicate a slight advantage in using this estimation function. However, statistical analysis reveals that based on this data, no definitive conclusions can be drawn on which estimation function is better.

Figure 5 presents the results for the shortest path application, for varying transmission ranges. The graphs are obtained by plotting the measured % error versus the normalized estimation function reported by the location discovery process for each shortest paths application. The plots correspond to preset communication ranges of 4, 6, 8, 10, 15, and 20 meters. Note that the horizontal axis, corresponding to the normalized location discovery estimation function, has a logarithmic scale. With this scale choice, the correlation between the measured application errors and the location errors become strikingly evident. As stated above, the application error here is measured by counting the number of shortest paths connecting arbitrary nodes determined

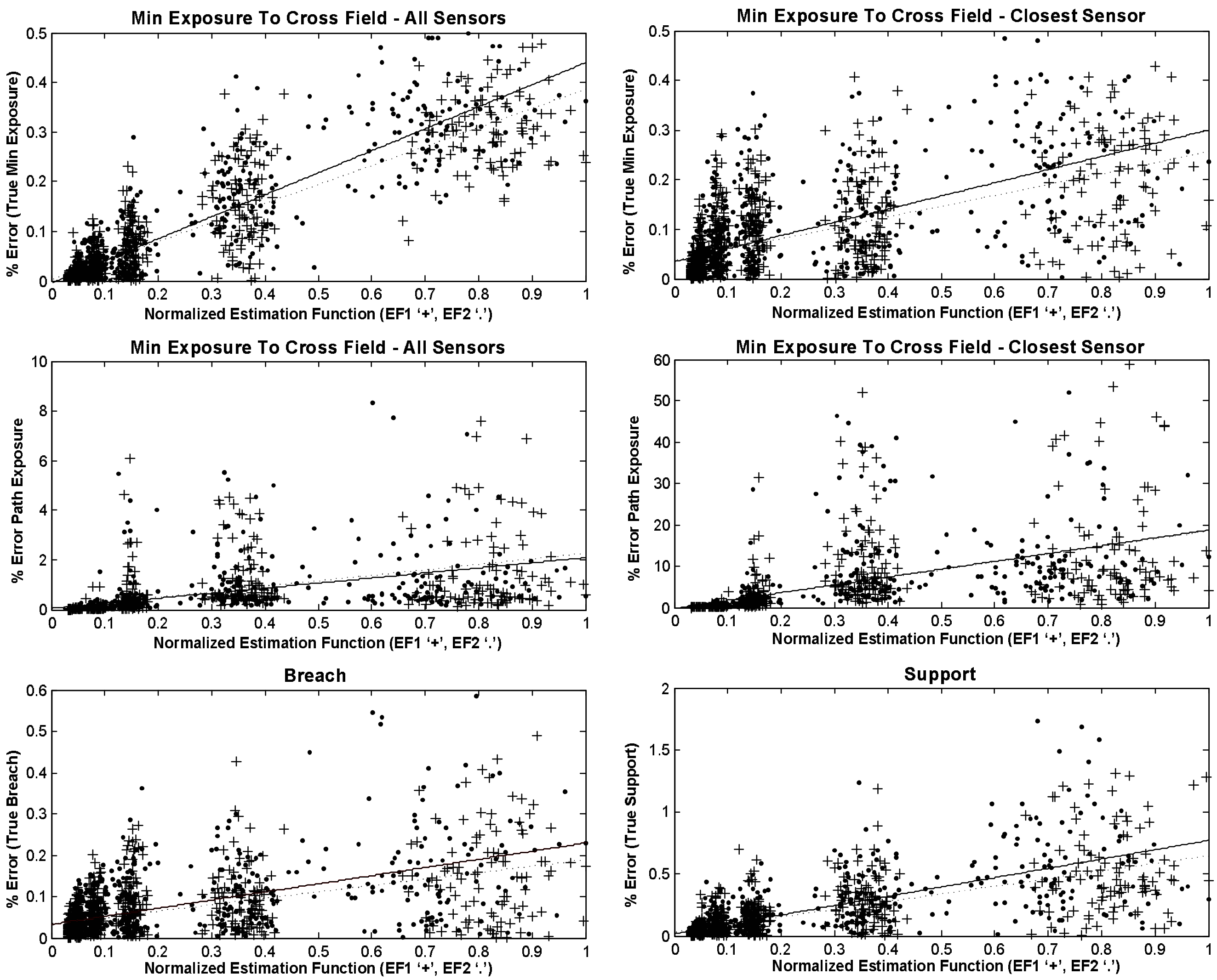

Figure 4: Error in min exposure, breach, and support vs. location discovery estimation function values. The top two plots show the min exposure error relative to the true min exposure (real locations). The middle row plots show the error in min exposure computed along the (faulty) path using faulty and real location information.

using faulty locations that are different from the paths obtained using the true locations. As in Figure 4, data points marked by a '.' correspond to values obtained using the maximum inconsistency (EF_2) and the data points marked by a '+' correspond to values obtained using the sum of inconsistencies (EF_1) estimation functions. It is interesting to note here that for medium communication ranges (i.e. 8m and 10m) the correlation between the estimation function value and the application error seem to be very different for relatively small objective function values compared to larger ones. While there seems to be a strong correlation between the errors for smaller values of estimation function, there seems to be almost no correlation for larger values. The communication range of 6m is most interesting since it is the most chaotic of all six cases. These trends confirm our intuition that for very short and very long communication ranges, location errors have a more predictable effect than for medium range communication ranges.

VII.B. Statistical Comparison

In the previous subsection, we studied the correlation between the location error estimation functions and application errors. In this subsection, we study how the information provided by the estimation functions can be useful. More specifically, we address this question: At what point are the results produced by the location discovery process useless to the application? The specific answer to this question will clearly be different for each application. However, in some instances such as breach, support, and minimal exposure, statistical profiles based on random network topologies can be used to obtain answers without relying on the costly and error prone location discovery process. To illustrate this point, Figure 6 shows histograms of statistical results obtained by considering 2000 instances of randomly deployed sensor networks in our 40m square field. For each case, the figures list the mean and the standard deviation of the obtained results. For example, we can expect that on average, the exposure along the minimal exposure path for

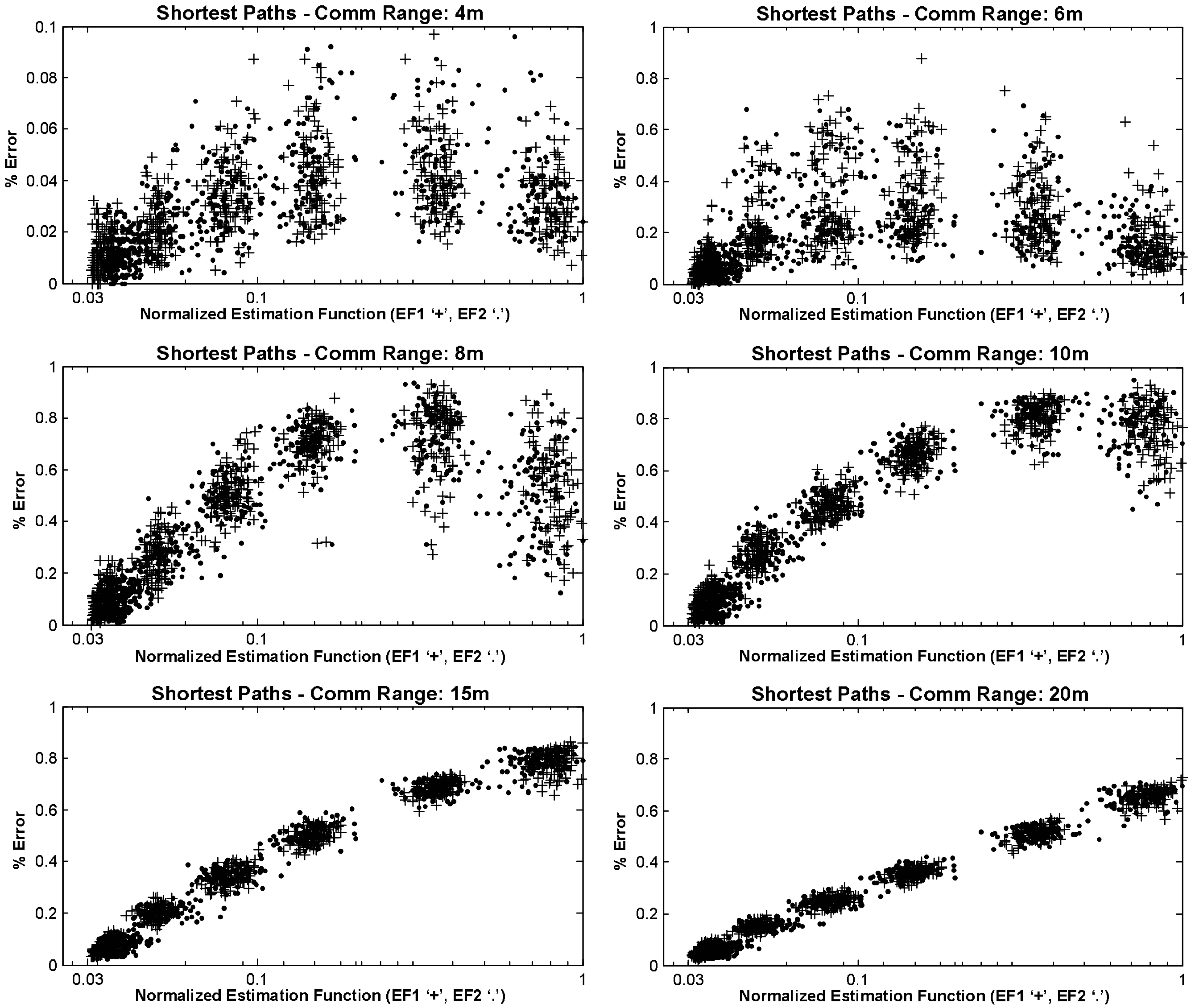

Figure 5: Shortest paths application error vs. the location discovery estimation function values at different preset communication range levels: 4m, 6m, 8m, 10m, 15m, and 20m.

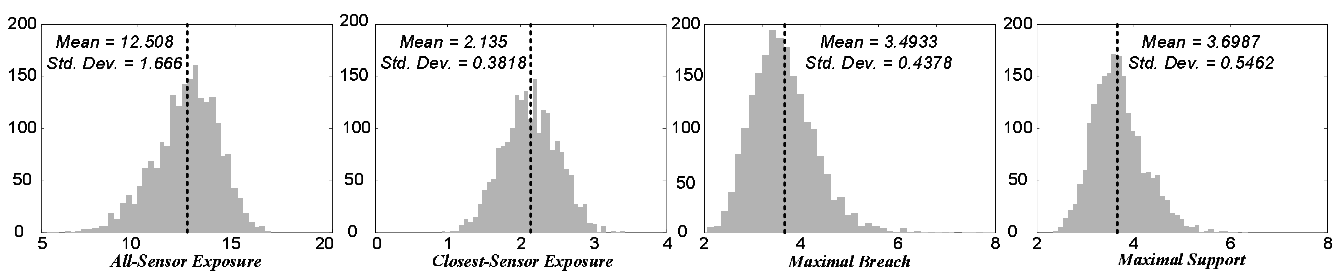

Figure 6: Statistics of minimal exposure, breach, and support for random 50-node sensor network topologies.

randomly deployed sensor networks with 50 nodes will be around 12.5. As we will see, in instances where location errors are large, this statistically obtained result based on random networks may be closer to the actual result than if we compute the minimal exposure path based on the faulty location information.

To compare these statistical results with our experimental data, consider the plots in Figure 7. Here, we consider 700 location discovery results for the same network of 50 nodes with varying average location errors. The horizontal axis for each plot represents the index of the data set. The vertical axis represent the actual computed value for all-sensor minimal exposure, closest-sensor minimal exposure, breach, and support using the faulty location data for

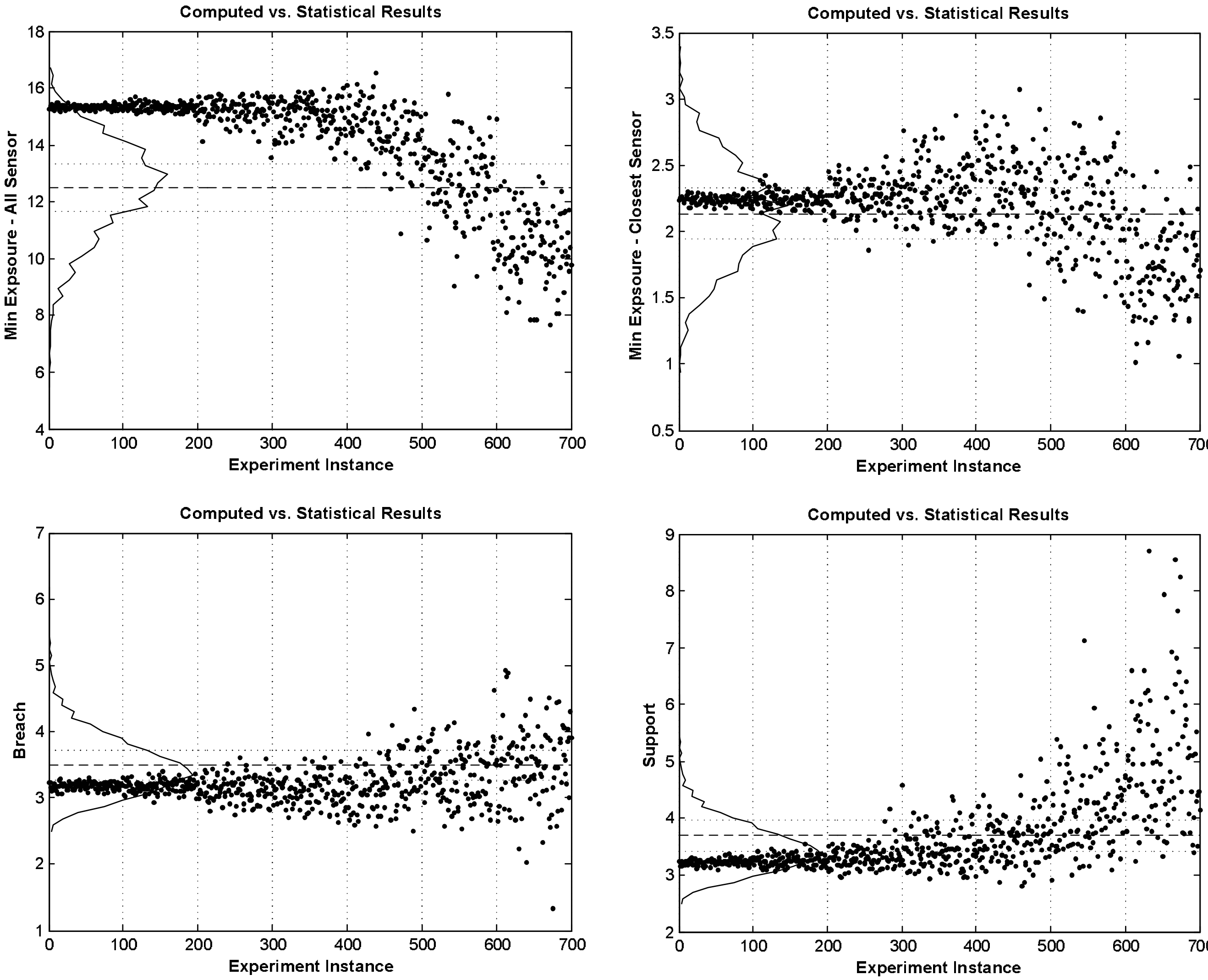

Figure 7: Computed (optimized) application data compared to statistics of random solutions (Fig 6). Horizontal axis indicates distance measurement errors ranging from 1% (0) to 40% (700) changed at the granularity indicated by vertical dashed lines.

each instance. The vertical grid lines show the granularity at which the initial location error levels were set that range between 1% to 40% respectively. The horizontal dashed- and dotted-lines indicate the mean and standard deviation corresponding to the statistical data. Note that the data in Figure 6 correspond to random 50-node network deployments while the results in Figure 7 are all from the same network instance. The variations in Figure 7 are purely the results of errors in node location. By simple inspection of these figures it is easy to see that roughly beyond the 500-th experimental instance, the computed results are less accurate on average than the statistically expected results. We can actually confirm this observation mathematically, since on average, the statistical mean (obtained from Figure) 6 is closer to the actual result than what we obtain through the optimization process. This implies that the measurement errors and errors inherent to the location optimization process are too large when propagated through the application level algorithms, starting at the 500-th instance (above 30% average location error). For example, in the case of the all-sensor-minimal-exposure application, statistical data indicate that we should expect an exposure level with mean 12.508 for a random 50 node network. Our specific network instance in Figure 7 has a minimal exposure level of roughly 15.25 (when the error is low). However, as the top-left graph in Figure 7 clearly indicates, as the errors increase along the horizontal axis, the statistical mean can predict the correct result with better accuracy on average than the results produced using the location discovery process.

VIII. Conclusion

We presented a discussion on location discovery errors in sensor networks and their effects on optimizations in WEASNs. We focused on location discovery induced errors in WEASNs since essentially the quality of service provided by WEASNs are dependent on accurate geographical node location information. In order to make our study tangible, we conducted statistical study of error sources, ways of propagations, and their impact on different applications. We described several different sources of error, whether and how different estimation functions capture different aspects of location errors, and then how the choice of a specific estimation function optimized during location discovery impacts the errors in results of different applications. Furthermore, we analyzed the impact of initial error measurements on the location discovery procedure and the effects on three fundamental yet distinct applications: exposure, worst- and best-case coverage (breach and support), and shortest paths (routing). We also demonstrated how such study of errors and their propagation can be used in the design and development of WEASN-specific algorithms. For example we showed how above a threshold error level (in distance measurements for example), existing statistical data can be used in computations instead of a costly optimization process which will essentially optimize noise and not useful data.

Acknowledgments

This material is based upon work partially supported by the National Science Foundation under Grant No. NI-0085773 and DARPA and Air Force Research Laboratory under Contract No. F30602-99-C-0128. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors' and do not necessarily reflect the views of the NSF, DARPA, or Air Force Research Laboratory.

References

[1] H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. F. Knight, R. Nagpal, E. Rauch, G. J. Sussman, and R. Weiss. Amorphous computing. *Communications of the ACM*, 43(5):74–82, May 2000.

[2] W. Adjie-Winoto, E. Schwartz, H. Balakrishnan, and J. Lilley. The design and implementation of an intentional naming system. *Operating Systems Review*, 33(5):186–201, Dec 1999.

[3] J. R. Agre, L. P. Clare, G. J. Pottie, and N. P. Romanov. Development platform for self-organizing wireless sensor networks. In *Proceedings of the SPIE - The International Society for Optical Engineering*, volume 3713, pages 257–68, Apr 1999.

[4] P. Bahl and V. N. Padmanabhan. Radar: an in-building RF-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000*, volume 2, pages 775–84, Apr 2000.

[5] G. Borriello and R. Want. Embedded computation meets the World Wide Web. *Communications of the ACM*, 43:59–66, May 2000.

[6] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low cost outdoor localization for very small devices. *IEEE Personal Communications, Special Issue on Smart Spaces and Environments*, 7(5):28–34, Oct 2000.

[7] D. Estrin, R. Govindan, and J. Heidemann. Embedding the Internet: Introduction. *Communications of the ACM*, 43(5):43, May 2000.

[8] L. Girod and D. Estrin. Robust range estimation using acoustic and multimodal sensing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2001)*, Maui, Hawaii, Oct 2001.

[9] A. Harter and A. Hopper. A distributed location system for the active office. *IEEE Network*, 8(1):62–70, Jan-Feb 1994.

[10] D. Kahaner, C. Moler, and S. Nash. *Numerical methods and software*. Prentice Hall, Englewood Cliffs, N.J., 1989.

[11] B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49(2):291–307, 1970.

[12] P. Krishnamurthy, K. Pahlavan, and J. Beneat. Radio propagation modeling for indoor geolocation applications. In *IEEE PIMRC'98*, Boston, Massachusetts, Sep 1998.

[13] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *IEEE Infocom 2001*, volume 3, pages 1380–1387, Apr 2001.

[14] S. Meguerdichian, F. Koushanfar, G. Qu, and M. Potkonjak. Exposure in wireless ad hoc sensor networks. In *Procs. of Int. Conf. on Mobile Computing and Networking (MobiCom '01)*, Jul 2001.

[15] S. Meguerdichian, S. Slijepcevic, V. Karayan, and M. Potkonjak. Localized algorithms in wireless ad-hoc networks: Location discovery and sensor exposure. In *MobiHOC 2001*, Oct 2001.

[16] D. Niculescu and B. Nath. Ad-hoc positioning system. Technical Report DCS-TR-435, Rutgers University, Apr 2001.

[17] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, May 2000.

[18] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *MobiCom 2000*, pages 32–43, 2000.

[19] J. M. Rabaey et al. PicoRadio supports ad-hoc ultra-low power wireless networking. *Computer*, 33(7):42–48, Jul 2000.

[20] A. Ralston and P. Rabinowitz. *A first course in numerical analysis*. McGraw-Hill, New York, second edition, 1978.

[21] T. S. Rappaport. *Wireless Communications - Principles and Practice*. Prentice-Hall, Inc., New Jersey, 1996.

[22] C. Savarese, J. M. Rabaey, and J. Beutel. Locationing in distributed ad-hoc wireless sensor networks. In *ICASSP 2001*, 2001.

[23] A. Savvides, C. Han, and M. Srivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *ACM Sigmobile (Mobicom)*, pages 166–179, 2001.

[24] M. J. Silventoinen and T. Rantalainen. Mobile station locating in GSM. In *IEEE Wireless Communication System Symposium*, Long Island, NY, Nov 1995.

[25] D. C. Steere et al. Research challenges in environmental observation and forecasting systems. In *MobiCom*, pages 292–299, 2000.

[26] G. Strang and K. Bore. *Linear Algebra, Geodesy, and GPS*. Wellesley-Cambridge, Wellesley, MA, 1997.

[27] G. S. Sukhatme and M. J. Mataric. Embedding robots into the Internet. *Communications of the ACM*, 43:67–73, May 2000.

[28] D. L. Tennenhouse and V. G. Bose. The SpectrumWare approach to wireless signal processing. *Wireless Networks*, 2(1):1–12, 1996.

[29] R. A. Thisted. *Elements of statistical computing*. Chapman and Hall, New York, 1988.

[30] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10(1):91–102, Jan 1992.

[31] M. Wylie and J. Holtzman. The non-line of sight problem in mobile location estimation. In *Proceedings of the 5th International Conference on Universal Personal Communications*, pages 827 – 831, Cambridge, MA, 1996.